

Visualizing State-Based Hypertension Progression Models

Amrita Gupta, Student Member, IEEE, Yu-Ying Liu, Student Member, IEEE, Jimeng Sun, Member, IEEE, and James M. Rehg, Member, IEEE

Abstract— We present a novel interactive visualization scheme for state-based hypertension progression modeling using hidden Markov models, applied to electronic health records of a cohort population enrolled in a hypertension management program. The visualization tool provides an interface for exploratory analysis and model validation, and improves the interpretability of the model results for healthcare researchers. We demonstrate a preliminary application of the visualization to compare states visited and transitions taken by two different subgroups with distinctive hypertension trajectories.

Index Terms—Disease progression modeling, continuous-time hidden Markov model, interactive visualization, medical informatics

1 INTRODUCTION

Hypertension is one of the main risk factors for cardiovascular disease and stroke, which claim over 17 million lives every year [1]. Several classes of drugs have been developed to help hypertensive patients maintain normal blood pressure, although the precise medication regimen needed to achieve blood pressure (BP) control for each individual is often difficult to determine. As a result, the medication regimen is often modified on a trial-and-error basis. Long term hypertension management programs [2, 3] can help participants achieve and maintain controlled blood pressure by encouraging regular home monitoring, facilitating medication adjustments and tracking disease parameters over time in electronic medical records.

Recently, there have been attempts to model hypertension from such electronic health record (EHR) data. One retrospective study [4] found that including EHR data improved the performance of regression models and gradient tree boosting for predicting cardiovascular-related death within 5 years, when compared to models trained with only traditional risk predictors. Several other studies were also performed to identify measures that act as risk predictors for the development of hypertension [5, 6]. However, for the most part these earlier works did not propose any framework for modeling the longitudinal progression of hypertension in diagnosed individuals.

More recent studies have begun to focus on the prediction of significant events within hypertension progression. Sun et al. used aggregated EHR data to predict transitions between controlled and uncontrolled hypertension, and also to identify which features produced the best prediction performance [7]. This approach, however, requires constant monitoring of patients' hypertension control status in order to determine the transition at the next clinical encounter. This highlights some of the difficulties of developing predictive models for hypertension given irregularly timed patient visits spanning only some segment of the disease progression.

In order to overcome some of these issues, our work models hypertension as a latent continuous-time Markov process evolving on a discrete state space, where each state is characterized by a certain range of systolic and diastolic blood pressures. This model structure can represent the continuous evolution of disease states using measurement data that arrives irregularly in time, and can be used to fill in missing measurements and predict future trajectories of

hypertension evolution. This class of model is referred to as a continuous time hidden Markov model (CT-HMM) [8-10]. It can also be used to study the specific effects of different medication regimens or individual-level features on hypertension evolution characteristics.

A basic challenge in using data-driven methods to improve healthcare is the development of appropriate visualization methods (viz). Viz enables health researchers to understand and validate a given model and perform exploratory data analysis, including the generation of hypotheses regarding disease progression. Without viz there is a danger that models that are fit to complex high dimensional datasets become “black box” analysis tools whose outputs cannot be explained or interpreted by health researchers.

This abstract presents FluxMap, the first interactive viz system for displaying and analyzing state-based disease progressions models based on the CT-HMM framework. FluxMap is designed to visualize state and transition properties of the CT-HMM and supports the identification and tracking of sub-groups of patients over time. There have been a few previous related works: Leiva-Murillo et al. used CT-HMM to model comorbidity of various diseases over time, illustrating their results using a static diagram of states and transitions [11]. However, only a small proportion of states and only the most dominant transitions can be included to avoid visual clutter. Another group used EventFlow software to look for common patterns in hypertensive drug usage and relate them to patient outcomes [12, 13]. FluxMap has the potential for use with any longitudinal state-based progression model, though here we present its use in conjunction with a CT-HMM model of hypertension built from longitudinal EHR data. To the best of our knowledge, this work is the first to demonstrate the application of an interactive visualization tool to state-based hypertension progression modeling.

2 METHODS

We analyze electronic health record data from a patient cohort involved in a hypertension management program. The dataset provides longitudinal data from several patients, each having a variable number of irregularly timed visits. At each visit, a number of observed variables are measured, though these measurements may be incomplete in the sense that not every feature is measured at every time point; for example, lab tests are not done at every visit. The hypertension model is based on a 2-dimensional discrete state space characterized by systolic and diastolic blood pressure values. Specifically, the discrete states are obtained by dividing the continuous space of each dimension into range bands. This discretization fits well with the typical practice of categorizing blood pressure using ranges of values. Also, the model allows flexible transitions between states without assumptions such as linear or polynomial changes usually used in regression model, and thus is suitable for exploratory data analysis.

- Amrita Gupta is with Georgia Institute of Technology. E-mail: agupta375@gatech.edu.
 - Yu-Ying Liu is with Georgia Institute of Technology. E-mail: yuyingliu@gatech.edu.
 - Jimeng Sun is with Georgia Institute of Technology. E-mail: jsun@cc.gatech.edu.
 - James M. Rehg is with Georgia Institute of Technology. E-mail: rehg@cc.gatech.edu.
- Proc. of IEEE VIS 2014 Workshop on Visualization of Electronic Health Records. Copyright retained by the authors.*

2.1 Continuous-Time Hidden Markov Model (CT-HMM)

We now provide a brief description of the CT-HMM [8][10] and its use in hypertensive progression modeling as follows. Each individual i in the subject population has n_i visits at irregularly spaced times (t_{i1}, \dots, t_{ini}) , where each visit has observed data (o_{i1}, \dots, o_{ini}) . These observations are generated conditionally on hidden states according to some emission probabilities $p(o/s)$, which are assumed to be distributed according to a multivariate Gaussian. The hidden states evolve as an unobserved Markov process, and the transitions between states are determined by transition intensities q_{rs} , which represent the instantaneous risk of transition from state r to s . From these, we can form an overall transition intensity matrix Q whose rows are made to sum to zero by setting the diagonal entries $q_{rr} = -\sum_s q_{rs}$. Then, for some time period t , the (r, s) entry of the transition probability matrix $P(t) = e^{tQ}$ describes the probability that an individual at state r at time t_0 will be in state s at time $t_0 + t$. The q_{rs} parameters are estimated using the Expectation-Maximization (EM) algorithm to maximize the likelihood of the data sequences. Specifically, we adopt the EM algorithm from [10] for efficient CT-HMM parameter learning in large state space setting.

The state structure for our hypertension progression model is a 2-dimensional discrete state space characterized by systolic and diastolic blood pressure range bands. Each state is allowed to instantaneously transition to its adjacent states, and a q_{rs} parameter is to be estimated between each pair of source and target state. An illustrative diagram is shown in Figure 1.

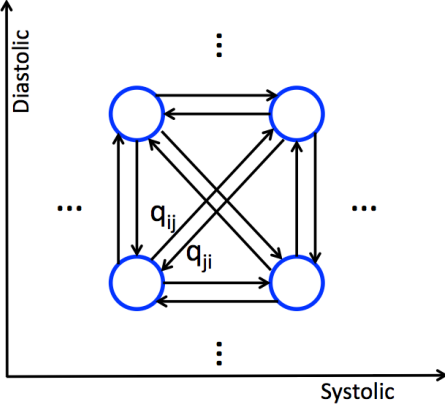


Fig. 1. The state and link structure for our hypertension progression model. The links represent allowed instantaneous transitions with transition intensity q_{ij} .

2.2 Interactive Web-based Visualization

In order to convey the results of the CT-HMM model in an easily interpretable way, we opted for an interactive web-based visualization scheme. The visualization is implemented in JavaScript with the support of D3.js and jQuery libraries. The states are arranged on a 2D grid layout according to the state's systolic and diastolic blood pressure ranges. Each state is represented as a circular node whose radius is proportional to the number of visits to that state. Transitions between states are bidirectional, so it was necessary to adopt an edge representation that encodes the direction of flow between the origin and destination nodes. This is important for being able to identify key behaviors in BP control like cycles or stable states. We used asymmetric Bezier curves to depict transitions between states, where the straight section of the curve is oriented towards the origin node and the more curved section points towards the destination node. Again, the width and opacity of each edge are proportional to the number of transitions that occur in the subject population, to emphasize high-magnitude transitions. This weighting

scheme was chosen over the scaling of edges based on transition intensities for several reasons. Using transition intensities could cause there to be links between every pair of states, however small, due to non-zero transition probabilities. The current weighting choice shows only the trajectories followed by subjects.

The second key component of the visual scheme is its interactivity. First, the user is allowed to interactively change the subject analysis population through filtering based on various features, such as medication class, BMI, risk level, lab test results, etc. Besides using these standard filters, the subset population may also be chosen based on some complementary analysis, such as clustering of subjects based on similar trajectories; such a use case is discussed in section 2.3. Once a subset of the patient population is chosen, their collective behavior over the state space can be inspected to identify major visited states and transitions. Clicking on a particular node highlights incoming and/or outgoing links with other states by fading out the other nodes and links slightly. If multiple patient subgroups are selected, the states are depicted as pie graphs to depict the proportion of measurements from each subgroup. This design is appropriate for use with a small number of patient subgroups as a means of identifying any states visited predominantly by one subgroup. The use of pie graphs

This tool thus enables users to examine the disease progression model in the context of a particular set of features of interest. For example, a user may wish to compare the effects of incorporating different drug classes into a hypertension treatment regimen on the amount of time spent in controlled-BP states versus uncontrolled-BP states. To do this, the user simply selects only subjects taking the medications of interest to be represented in the graph, displaying the states visited and the proportions of each medication class at each one. Hence this visualization scheme combines an intuitive presentation of the CT-HMM model predictive results with the capacity for some visual exploratory analysis.

2.3 Application: Trajectory Clustering

To demonstrate the exploratory visual analysis enabled by our tool, we first trained the CT-HMM on the dataset and produced decoded state sequences for all subjects for the window of time they were observed. The resulting sequences of states were compared between subjects to produce a similarity score based on the weighted Hamming distance. More specifically, the "distance" between two states is taken as the number of jumps or edits needed to move from one state to the other. To compare two sequences, first an alignment is chosen such that the sequences overlap by at least 3 states; then, the cumulative distance between overlapping states is found and normalized. Different alignments are tested by sliding one sequence along the other, and the final similarity score is chosen as the minimum average distance over all possible alignments. This approach was adopted to account for the fact that subjects' participation in the hypertension management program did not coincide with any distinctive time point in the progression of their hypertension. Finally, the subjects' trajectories are grouped via hierarchical clustering to obtain subsets with similar state sequences.

3 RESULTS

The process described in section 2.3 produced a hierarchical clustering of subjects, with two distinct low-level clusters depicted in Figures 2 and 3. Cluster 1 contains 194 subjects with lower diastolic BP on average compared to the 282 subjects in cluster 2. Further, it appears that subjects in cluster 1 often show changes in systolic pressure without accompanying changes in diastolic pressure, as evidenced by the slightly larger weights and numbers of horizontal transitions compared to vertical or diagonal ones. In contrast, subjects in cluster 2 have a noticeably higher tendency for coupled changes in systolic and diastolic BP, as shown by the prominent diagonal transitions in Figure 3.

These differences are evident in the combined view using pie charts shown in Figure 4. By limiting the number of classes to just 2, it is easy to see from the mostly dark brown nodes with diastolic BP 55 that these states are nearly exclusively visited by subjects in cluster 1. Similarly, states with diastolic BP 85 and higher are frequented by subjects in cluster 2 more than cluster 1.

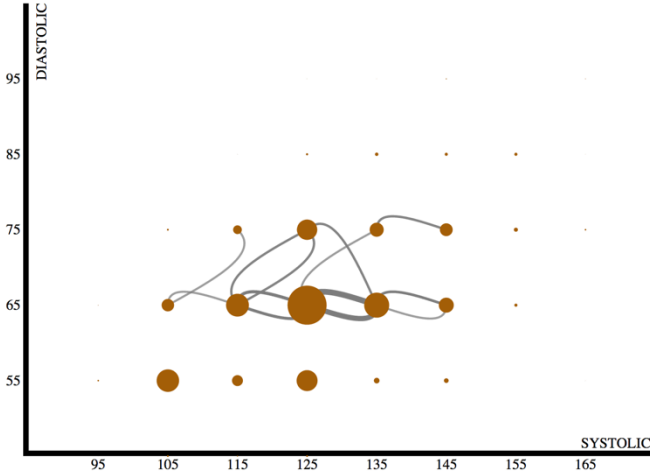


Fig. 2. States and transitions by subjects in cluster 1.

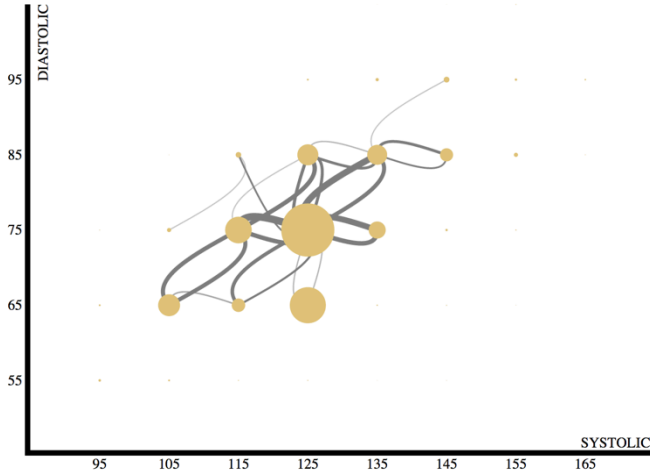


Fig. 3. States and transitions by subjects in cluster 2.

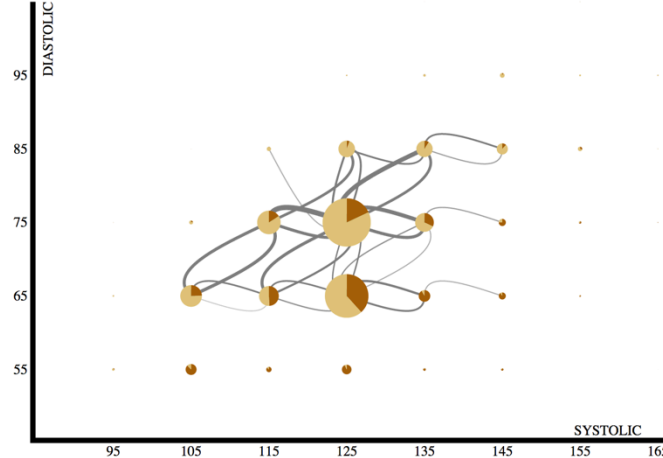


Fig. 4. Clusters 1 and 2 shown together.

4 CONCLUSION

In this paper, we presented an interactive visualization system for disease progression modeling using CT-HMM, with the potential for use as a validation tool for the model as well as an interface for conducting exploratory analysis. We demonstrated one use case of the tool for viewing states and trajectories; the case compared subjects from two different subgroups of hypertensive patients obtained by a hierarchical clustering of state sequences decoded by the CT-HMM. This framework can be extended to examine more elaborate questions related to the effects of different features on the trajectories taken by subjects. This may be useful for identifying phenotypes or different behaviors within hypertension management.

The current visualization scheme has several possibilities for further development. At present, the identification of true transition sequences (with 2 or more transitions) is not possible due to the aggregation of transitions across subjects. While the current approach provides a summary overview of major transitions, it is necessary to include some mechanism for recognizing and choosing specific sequences of states that are visited as part of a major path or cycle. These capabilities will become possible as the visualization scheme matures. Another possible limitation of the FluxMap visualization is its limitation to a 2-dimensional state space, whereas the CT-HMM model for which it was developed can support a higher-dimensional state space. This disparity can be partly addressed by adopting a side-by-side paneled view, in which each panel shows a 2D projection or collapse of the ND state space. However, this drastically reduces the interpretability of the visualization scheme and is untenable for more than 4 dimensions.

REFERENCES

- [1] World Health Organization. "A global brief on hypertension: silent killer, global public health crisis: World Health Day 2013." (2013).
- [2] Mehos, Brenda M., Joseph J. Saseen, and Eric J. MacLaughlin. "Effect of pharmacist intervention and initiation of home blood pressure monitoring in patients with uncontrolled hypertension." *Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy* 20.11 (2000): 1384-1389.
- [3] Jaffe, Marc G., et al. "Improved blood pressure control associated with a large-scale hypertension program." *JAMA* 310.7 (2013): 699-705.
- [4] Kennedy, Edward H., et al. "Improved cardiovascular risk prediction using nonparametric regression and electronic health record data." *Medical care* 51.3 (2013): 251-258.
- [5] Lukas, Arno, et al. "Body mass index is the main risk factor for arterial hypertension in young subjects without major comorbidity." *European journal of clinical investigation* 33.3 (2003): 223-230.
- [6] Bozorgmanesh, Mohammadreza, et al. "A point-score system superior to blood pressure measures alone for predicting incident hypertension: Tehran Lipid and Glucose Study." *Journal of hypertension* 29.8 (2011): 1486-1493.
- [7] Sun, Jimeng, et al. "Predicting changes in hypertension control using electronic health records from a chronic disease management program." *Journal of the American Medical Informatics Association* 21.2 (2014): 337-344.
- [8] Jackson, Christopher H. "Multi-state models for panel data: the msm package for R." *Journal of Statistical Software* 38.8 (2011): 1-29.
- [9] Bartolomeo, Nicola, Paolo Trerotoli, and Gabriella Serio. "Progression of liver cirrhosis to HCC: an application of hidden Markov model." *BMC medical research methodology* 11.1 (2011): 38.
- [10] Liu, Yu-Ying, et al. "Longitudinal Modeling of Glaucoma Progression Using 2-Dimensional Continuous-Time Hidden Markov Model." *Medical Image Computing and Computer-Assisted Intervention-MICCAI 2013*. Springer Berlin Heidelberg, 2013. 444-451.
- [11] Leiva-Murillo, Jose, Antonio Artés-Rodríguez, and Enrique Baca-García. "Visualization and prediction of disease interactions with continuous-time hidden markov models." (2011).

- [12] Bjarnadottir, Margret, and Sophia Wu. "WORKSHOP: Exploring Temporal Patterns in Hypertensive Drug Therapy Thursday, May 29, 2014." *HCIL*. N.p., n.d. Web. 31 Aug. 2014.
- [13] Rind, A., et al. Interactive information visualization for exploring and querying electronic health records: A systematic review. Vol. 19. Technical Report HCIL-2010, 2010.